Notes on Problem Set Solutions

Question 1.

For this question, we should count number of employees who response positively to the policy variables. We can use different measurement to catch employees' attitude toward them. First, we can count the number of people who say "very likely". We can also count the number of the who say "likely". We can even calculate percentage of people who response positively. We draw our conclusions based on the numbers and percentages. Second, notice that the response is ordered, with "0" stands for "very likely", with "1" stands for "likely" and "2" stands for not likely. We can actually average the number for each policy. The lower the averaged value, the more positive the response is. Most of you use the first approach. The query used for the second approach is presented below.

You may be familiar with query design view and hate the Sql script. To know what they look like in the query design view, you can copy them to your Access Sql window and then change to design view.

a. Convert data type of policy variable from text to number, and then run the following queries.

```
SELECT Count(PUB_DISC)/Count(surveyno) AS Expr1, Avg(EMPLOYEE.PUB_FLEX) AS AvgOfPUB_FLEX, Avg(EMPLOYEE.PUB_DAYCAR) AS AvgOfPUB_DAYCAR, Avg(EMPLOYEE.PUB_EMGNCY) AS AvgOfPUB_EMGNCY, Avg(EMPLOYEE.PUB_PKRATE) AS AvgOfPUB_PKRATE, Avg(EMPLOYEE.PUB_INFO) AS AvgOfPUB_INFO, Avg(EMPLOYEE.PUB_DISC) AS AvgOfPUB_DISC
FROM EMPLOYEE
WHERE (((EMPLOYEE.MODE1_TO) In ("1","2","3")));
```

With this query, we find policy associated with lowest value are discount passes and emergency rider services

b. for the four greatest four agencies, first find out them by the following query

```
SELECT Count(employee.SURVEYNO) AS CountOfSURVEYNO, employee.DOT_BRANCH
FROM employee
GROUP BY employee.DOT_BRANCH
ORDER BY Count(employee.SURVEYNO) DESC;
```

They are FAA, USCG, OST, FHWA

Then find out the most positively responded policies in these four agencies

```
SELECT Avg(EMPLOYEE.PUB_FLEX) AS AvgOfPUB_FLEX, Avg(EMPLOYEE.PUB_DAYCAR) AS AvgOfPUB_DAYCAR, Avg(EMPLOYEE.PUB_EMGNCY) AS AvgOfPUB_EMGNCY, Avg(EMPLOYEE.PUB_PKRATE) AS AvgOfPUB_PKRATE, Avg(EMPLOYEE.PUB_INFO) AS AvgOfPUB_INFO, Avg(EMPLOYEE.PUB_DISC) AS AvgOfPUB_DISC
FROM EMPLOYEE
WHERE (((EMPLOYEE.DOT_BRANCH) In ("FAA","USCG","OST","FHWA")) AND (((EMPLOYEE.MODE1_TO) In ("1","2","3"))));
```

The most positively responded policies are the same as above.

To make recommendation branch by branch, we can use the following query.

```
SELECT EMPLOYEE.DOT_BRANCH, Avg(EMPLOYEE.PUB_FLEX) AS AvgOfPUB_FLEX, Avg(EMPLOYEE.PUB_DAYCAR) AS AvgOfPUB_DAYCAR, Avg(EMPLOYEE.PUB_EMGNCY) AS AvgOfPUB_EMGNCY, Avg(EMPLOYEE.PUB_PKRATE) AS AvgOfPUB_PKRATE, Avg(EMPLOYEE.PUB_INFO) AS AvgOfPUB_INFO, Avg(EMPLOYEE.PUB_DISC) AS AvgOfPUB_DISC
FROM EMPLOYEE
WHERE (((EMPLOYEE.MODE1_TO) In ("1","2","3")));
```

GROUP BY EMPLOYEE.DOT_BRANCH
HAVING (((EMPLOYEE.DOT_BRANCH) In ("FAA","USCG","OST","FHWA")));

Almost all of you are OK with this questions. Some of you may understand the question a bit differently, it is OK with me when I graded it. You are given full credit or be fined only slightly. The main point is that you run the appropriate queries and get the important numbers. However, several students don't show the needed numbers to make the recommendation.

Question 2.

A. Table of alphabetical listing by COUNTY and MODE1 (name label) and Avg of T_COMMUTE:

The query design looks like this:
The SQL looks like this:
You may use commuting distance instead of commuting time. Both are OK. The table should look something like this:

<table>
<thead>
<tr>
<th>Field</th>
<th>Table</th>
<th>Total</th>
<th>Sort</th>
<th>Show</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>COUNTY</td>
<td>employee</td>
<td></td>
<td>Group By</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>dcomodpc</td>
<td>employee</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Avg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ascending</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fairfax, Montgomery, P.G., Wash DC</td>
</tr>
</tbody>
</table>
B. The first and second most popular modes of travel in Montg are car pool and metro rail

For P G, the most popular ones are car pool and drive alone

c. In Montg, 255 people have secondary commuting mode
SELECT Count(employee.SURVEYNO) AS CountOfSURVEYNO
FROM employee
WHERE (((employee.COUNTY)="Montg") AND ((employee.MODE2_TO)<"0");

For those without secondary commuting mode, the most popular primary commuting mode is carpool (177 people carpool).

SELECT dotmodpc.MODE_1, employee.COUNTY, Count(employee.SURVEYNO) AS CountOfSURVEYNO, Avg(employee.D_COMMUT) AS AvgOfD_COMMUT, employee.MODE2_TO
FROM dotmodpc INNER JOIN employee ON dotmodpc.MODE1_CODE = employee.MODE1_TO
GROUP BY dotmodpc.MODE_1, employee.COUNTY, employee.MODE2_TO
HAVING (((employee.COUNTY)="Montg") AND ((employee.MODE2_TO)="0"))
ORDER BY employee.COUNTY, Count(employee.SURVEYNO) DESC;

Question 3.

A. County block groups with highest percentages of government workers:

<table>
<thead>
<tr>
<th>Counties</th>
<th>Tract:</th>
<th>Block group</th>
<th>Percentage of Government Workers:</th>
<th>Block group key</th>
</tr>
</thead>
<tbody>
<tr>
<td>25013</td>
<td>Hampden</td>
<td>8114</td>
<td>100%</td>
<td>250138114003</td>
</tr>
<tr>
<td>25015</td>
<td>Hampshire</td>
<td>8203</td>
<td>63.6%</td>
<td>250158203004</td>
</tr>
<tr>
<td>25027</td>
<td>Worcester</td>
<td>7384</td>
<td>71.1%</td>
<td>250277384009</td>
</tr>
</tbody>
</table>

Note that full credit was given to those who had the answers in the table above. Government workers were determined by adding P0790003, P0790004, P0790005. The percentages were determined using the universe of employed persons as derived by adding up the fields P0790001, P0790002, P0790003, P0790004, P0790005, P0790006, P0790007. As seen in the Census description below, adding up the like categories provides an appropriate universe. There was confusion among some students who calculated percentages with the same universe used in Lab E. That is, the universe determined by adding male and females who worked in 1989. In the Census, work status has a different definition than employment status. While it was appropriate to use males and females who worked as the universe in the Lab E calculations, it produces wrong results when used with class of worker which is based on the universe of employed persons 16 years and over. See below.

P79. CLASS OF WORKER(7) [7]
Universe: Employed persons 16 years and over:
Private for profit wage and salary workers: P0790001
Private not for profit wage and salary workers: P0790002
Local government workers: P0790003
State government workers: P0790004
Federal government workers: P0790005
Self employed workers: P0790006
Unpaid family workers: P0790007

Because the confusion was due to these subtle but important differences, we gave credit to answers based on the universe of workers that was used in Lab E. However, full credit was given to the correct answers.

Question 4.

Many students did well on this question and correctly chose Map 2 as the most important map to use in the analysis. Note that Map 2 is the most appropriate map because it shows the relative density of adults without degrees. Normalizing the data (e.g., determining persons per square kilometer) provides a standard way to compare regions. However, the argument on highest density alone is not enough. We should also think of the density of the neighbors. Think of whether it is easily accessible from other towns. That's a very important reason we use maps instead of using only the underlying tables.